CHAPTER IV

DISCUSSION

4.1 Genetic diversity and population differentiation of *A. cerana* in Thailand

Sequencing of the amplified lrRNA gene of *A. cerana* indicated low levels of genetic diversity between bees originating from different geographic locations in Thailand. The percent nucleotide divergence of the lrRNA gene was 0.15% - 1.70% whereas that between *A. cerana* and a representative *A. mellifera* was 13.59%.

Smith and Hagen (1997) sequenced the non-coding intergenic region of COI-COII (68-73 bases) of 110 *A. cerana* individuals. A total of 35 point mutations was observed. This indicated that this region is highly polymorphic and very useful in determining intraspecific biogeography of *A. cerana* over most of its geographic distribution. A lower level of polymorphism was observed in lrRNA gene sequences (57 point mutations from 653-654 nucleotides). Although the number of bees used for DNA sequencing analysis was rather small, variability of *Dra* I recognition sites of bees from different geographic origins were found. Therefore, genetic differentiation analysis of *A. cerana* was simplified to RFLP of the lrRNA gene with a single restriction enzymes, *Dra* I.

Distribution and frequencies of *Dra* I-digested lrRNA haplotypes indicated clear differentiation between north-to-central (north, north-east and the central region) and southern latitude (peninsular Thailand, Phuket and Samui Islands) bees. The common haplotypes A and B were not concurrently found in any geographic sample of *A. cerana* in Thailand. The variant haplotype D was not found at the contact zone indicating restricted female gene flow between bees from these regions.

Other analyzes carried out in this laboratory using polymorphism of an ATPase 6-8 and AT-rich region of the same sample set did not reveal shared genotypes between bees from the north-to-central region and peninsular Thailand (Songram, 1997; Pootong, 1998). This suggests that bees from north-to-central and

southern regions may have been colonized by two independent ancestral populations, one from the north, one from the south.

The C haplotype was specifically found with a high frequency in the Samui sample but was absent from others implying that this haplotype was recently derived from the B genotype. Genetic drift may have played an important role influencing the high level of the haplotype C in Samui Island. Sylvester *et al.* (1998) reported the distinctness of *A. cerana* from peninsular Thailand and Samui Island based on morphometric analysis and concluded that these samples have evolved separately for a period of time.

Geographic heterogeneity tests and F_{ST} statistics indicated that the Thai *A. cerana* population is not panmictic but fragmented into several groups ($P < 0.0001$). Significant genetic differentiation between each of the north-to-central and southern samples ($F_{ST} = 0.618 - 0.982$, $P < 0.0001$) were observed. Within the southern region, significant genetic difference was observed between peninsular Thailand and Samui ($P < 0.0001$) but not between Phuket and Samui Islands located at different coastal sides ($P > 0.05$). This unexpected circumstance could have resulted from a small sample size of the Phuket sample ($n = 7$) in this study.

Results on genetic differentiation of *A. cerana* in the southern region suggest that the arrival of *A. cerana* to the Samui Island probably occurred prior to the most recent rise of sea level of this region (approximately 5000 years before present at 5 metres above the present sea level) (Dall *et al.*, 1990; Pianka, 1994).

Analysis of molecular variance (AMOVA) and other analyses indicated that six geographic samples of *A. cerana* in Thailand should be divided to 3 groups (populations); A) north, north-east and the central region, B) peninsular Thailand and Phuket Island and C) Samui Island.

Although analysis of genetic diversity of various organisms by DNA sequencing is the most direct and reliable technique, we demonstrated that restriction analysis of lrRNA gene polymorphism using *Dra* I was equally potential and sufficient to examine genetic differentiation of *A. cerana* in Thailand. Nevertheless, patterns of distributions and genetic diversity levels of *A. cerana* should be further

studies covering vast geographic areas of all subspecies using more number of genes and/or restriction enzymes.

4.2 Isolation and characterization of AcMRJPs

Although population genetic analysis indicated the existence of 3 different populations of *A. cerana* in Thailand, bees from different geographic origin did not revealed significant different ability to produce RJ ($P > 0.05$). Crude RJ from different *A. cerana* origins analyzed by Coomassie brilliant blue R-250 stained SDS-PAGE showed relatively identical ratio of major 50 kDa and 80 kDa proteins.

Difference patterns of major bands of crude RJ of *A. cerana* (50 kDa, 80 kDa and 80kDa) and *A. mellifera* (55 kDa, 60 kDa, 70 kDa and 80 kDa) were observed. Previously, Takenaka and Takenaka, 1996 compared electrophoretic patterns of RJ of these species using SDS page analysis and indicated that 6 bands (Nos. 4, 6, 7, 12, 16 and 21 from large to small sizes in order)were shared where 4 of these (Nos. 6, 7, 12 and 16) were heavier stained in *A. cerana* RJ. The most intense stained band in RJ of *A. mellifera* was a 60 kDa band whereas that of *A. cerana* was a 50 kDa band. The abundant of these bands in RJ were concordant to results previously reported (Takenaka and Takenaka, 1996).

A DEAE cellulose column chromatography was commonly used to purify *A. mellifera* RJ. Three protein peaks were purified and eluted out at 0.11 - 0.16 M, 0.16 - 0.22 M and 0.22 - 0.3 M NaCl in 20 mM Tris-HCl buffer pH 7.5 and 1 mM EDTA (Schmitzova *et al.*, 1998) or at 0.05 M, 0.10 M and 0.20 M NaCl in 20 mM Tris-HCl pH 7.2 (Tomoda *et al.*, 1977). In contrast, *A. cerana* RJ isolated by a DEAE cellulose column chromatography revealed only two major protein peaks (DEAEP1 and DEAEP2) eluted at 0.05 M and 0.15 M NaCl in 20 mM Tris-HCl buffer pH 7.5 containing 1 mM EDTA. From SDS-PAGE, the protein peak DEAEP1 showing the high content of a 55 kDa protein and very low content of a 80 kDa protein while protein peak DEAEP2 showing a major protein with molecular weight of 50 kDa.

Using DEAE ion-exchanged chromatography, a 80 kDa protein (subsequently verified to be AcMRJP3) could not be purified and used for further characterization. Albert *et al.* (1999b) reported the existence of NQNA(D/N/G) pentapeptide units at the *N*-terminus and (K/R)QN(D/G)N pentapeptide at the C terminus of ACMRJP3.

Each unit of the latter repeat started with Lys or Arg (positively charges). Therefore, this region may have interfered purification of AcMRJP3 RJ as similar repeated units with more positively charges (RRNGN and KQNDN) were found.

The purification process was altered to a Q-Sepharose column chromatography instead. Three protein peaks (A, B and C) were eluted by linear NaCl gradient at 0.075 M, 0.175 M and 0.25 M NaCl in 20 mM Tris-HCl buffer pH 7.5 and 1 mM EDTA, respectively. Peak A contained two bands of 80 kDa and 55 kDa after SDS-PAGE while both of peak B and C showed only one major band at 50 kDa. The major protein bands of proteins peak DEAEP1 and DEAEP2 was *N*-terminal amino acid sequenced and then verified to be MRJP2 and MRJP1 homologues, respectively.

Each peak (A, B, C) from Q-Sepharose column was further purified using a Sephadex G-200 column. Protein in peak A was separated into two peaks (A1 and A2). The protein A1 showed 115 kDa and 80 kDa in native and denatured molecular weight while a protein A2 showed 55 kDa in both forms. Both protein peak B and C showed a single peak on Sephadex G-200, called B1 and C1 with the native molecular weight of 50 kDa and 300 kDa, respectively. SDS-PAGE analysis indicated identical major protein bands between B and B1 and between C and C1. Therefore, only a single purification step of RJ of *A. cerana* through a Q-sepharose column may be enough to purify protein peak B1 and C1 (subsequently clarified to be AcMRJP1). Using SDS-PAGE analysis, B1 was a monomeric protein whereas C1 was an oligomeric protein containing identical subunits.

For further characterization of purified RJ proteins, *N*-terminal amino acid sequences of different purified proteins were compared to AmMRJPs described by Schmitzova *et al.* (1998). Proteins A1, A2, B1 and C1 were identified to be AmMRJP3, AmMRJP2, monomeric AmMRJP1 and oligomeric AmMRJP1 homologues, respectively. Simuth (2001) separated RJ of A. mellifera by ultracentrifugation and reported that denatured and native molecular weight of AmMRJP1 was 55 kDa and 420 kDa, respectively. Sizes of both denatured and native AcRJP1 were smaller than those of AmMRJP1 which were possible due to different sizes of carbohydrate side chains and/or subunit polymorphism.

Proteins A1 and A2 were classified to be AcMRJP3 and AcMRJP2 and first reported by this study. The molecular weight of the former was 80 kDa for denatured and 115 kDa for native forms whereas the latter exhibited sizes of 55 kDa for both denatured and native forms. Putative molecular weight of AcMRJP1, AcMRJP2 and AcMRJP3 calculated from amino acid sequences was 48.8 kDa, 52.4 kDa and 69.5 kDa, repectively. Size differences between inferred and monomeric proteins purified from RJ were possibly explained by the existence of oligosaccharides in these MRJPs.

PAS staining and *N*-glycosidase F digestion revealed that all purifed AcMRJPs were glycoproteins. Results were concordant to the existence of putative *N*-glycosylation suites in these proteins. Kimura, Washino and Yonekura (1995) and Kimura *et al.* (1996) identified the structure of *N*-linked glycan in AmMRJP1 and AmMRJP2. More recently, Kimura *et al.* (2000) reported highly enriched mannose, accounting for 71.6 % of total *N*-glycan, in crude RJ proteins

Using isoelectric focusing, all purified AcMRJPs of *A. cerana* purified proteins were characterized. The pI values indicated that AcMRJP1 was an acidic protein whereas AcMRJP2 and AcMRJP3 were basic proteins. Results from the present study were not contradictory to pI values of AmMRJP1 and AmMRJP2 previously reported by Hane and Simuth (1992) and Bilikova *et al.* (1999). All purified AcMRJPs with the exception of the protein C1 (oligomeric AcMRJP1) showed variant types of the same protein family showing different pI values. This indicated polymorphism of proteins B1 (monomeric AcMRJP1), A2 (AcMRJP2) and A1 (MRJP3).

Polymorphism revealed by different pI values may be caused by charged amino acid (K or R and D or G) differences between different variants. In this study, the repeated unit of NQ (K/N) N (N/T) were found in the deduced AcMRJP2 while deduced AcmRJP3 contained two repeated units; NQN (A/D) (N/T) and (R/K/N) (Q/R) N(G/D/A/S) N at the C-terminus.

In this study, degradation of AcMRJPs was investigated under various conditions (at –20 °C to 37 °C for 0 to 30 days). AcMRJP2 and AcMRJP1 were highly stable protein whereas AcMRJP3 was not stable and degraded at both storage temperature of 4 °C and 37 °C. It should be noted that monomeric AcMRJP1 (protein

peak B1) was more stable than oligomeric AcMRJP1. Degradation of oligomeric AcMRJP1 (protein peak C1) with the appearance of a 48 kDa band was found after stored at 37 °C for 3-15 days.

Kamakura *et al* (2001) reported effects of the temperature on stability of AmMRJPs during storage period. RJ of *A. mellifera* was tested for the stability at various conditions (at 4°C to 50 °C for 0 to 7 days). After that, RJ was analyzed by SDS-PAGE, native PAGE and HPLC. Degradation of AcMRJP1 was directly proportional to both storage temperature and period. The result suggested that AmMRJP1 could be used as a marker for freshness of RJ.

The amount of individual AmMRJP family in total RJ protein was previously determined after SDS-PAGE separation. The relative content of AmMRJP1, AmMRJP2, AmMRJP3, and AmMRJP5 was 31 %, 16 %, 26 % and 9 %, respectively (Schmitzova *et al.*, 19998). Different ratio of total protein A1 (AcMRJP3): A2 (AcMRJP2): B1 (monomeric AcMRJP1): C1 (oligomeric AcMRJP1) was found among three *A. cerena* populations in Thailand (north to central, peninsular and Samui Island). Total AcMRJP1 was the most abundant (64.59 % - 67.15% of total RJ protein families) while AcMRJP2 and AcMRJP3 were accounted for 22.95 % - 28.31 % and 5.20 % - 12.32 %, respectively. The average ratio of AcMRJP1 : AcMRJP2 : AcMRJP3 content was 7.2 : 2.7 : 1.

In this study, the main members of protein families in RJ could be purified. These proteins represent a balanced mixture for nourishing both honeybee larvae and queen. Although the nutritional function of all purified protein was proposed, different properties of purified protein were found. Other biological roles of these proteins should be studied. Moreover, This result suggested that RJ purified proteins should be stored below –20 °C after harvested for their stability.

4.3 Molecular cloning and characterization of AcMRJP cDNAs

ESTs established from hypopharyngeal mRNA of *A. cerana* nurse bees analyzed by this study contained 50.0 % AcMRJP1, 6.06 % AcMRJP2, 6.06 % AcMRJP3 and 1.52 % AcMRJP4, respectively. Nevertheless, AcMRJP5 was not found when RJ of *A. cerana* was directly purified. AcMRJP5 was not found both from the cDNA library and RJ purification.

The complete nucleotide sequence of AcMRJP1 was derived from three nucleotide sequences of cDNA library clones while the complete nucleotide sequences of AcMRJP2 and AcMRJP3 were obtained from RT-PCR cloning. Nucleotide sequences of AcMRJP1, AcMRJP2 and AcMRJP3 showed high homology with those of AmMRJP1 (93.1 %), AmMRJP2 (92.10 %) and AmMRJP3 (91.5 %), respectively. A repeated unit was found at the *C*-terminus of AcMRJP3 coding region therefore polymorphism of AcMRJP3 may have been occurred as previously reported in AmMRJP3 (Albert *et al.*, 1999b). Polymorphism of AcMRJP3 can be studied by PCR using a pair of primers flanked of the repeated unit. Length polymorphism of the amplified product should be found. In addition, a repeated unit at the *C*-terminus of AcMRJP2 was also observed. Variants of AcMRJP2 should have occurred.

Identical number of putative glycosylation were observed between MRJP1 and MRJP2 of *A. cerana* and *A. mellifera* (3 and 2 *N*-linked glycosylation sites, respectively) (Schmitzova *et al.*, 1998). However, putative *N*-linked glycosylaton site of AcMRJP3 was 5 times greater than that in AmMRJP3. The *O*-liked glycosylation site was not found in any AcMRJPs. The predicted pI values of AcMRJP1, AcMRJP2 and AcMRJP3 were 5.5, 8.02, 8.8, respectively. These were roughly comparable to values determined by IEF analysis of purified AcMRJPs described previously.

The essential amino acid content of AcMRJP1, AcMRJP2 and AcMRJP3 was relatively high (48.3 %, 46.4% and 39.5 %, respectively) which is comparable to that of AmMRJP1 (48 %), AmMRJP2 (47 %), AmMRJP3 (39.3 %) (Schmitzova *et al.*, 1998).

The average sequence divergence between the same MRJP family from different bee species (e.g. MRJP1 of *A. cerana* and *A. mellifera*) was lower than that

between divergence calculated from pairs of different protein family within the same species (e.g. comparisons between AmMRJP1-AmMRJP2, AcMRJP1-AcMRJP2 etc.). The bootstrapped phylogenetic tree revealed closer relationships between the same RJ protein families. The phylogenetic trees constructed from nucleotide and amino acid sequence divergence supported the occurrence of MRJPs through gene duplication and an earlier separation of MRJP4. The occurrence of MRJP1, MRJP2 and MRJP3 were possibly existent at approximately the same evolutionary time whereas more data about MRJP5 are required for unambiguous conclusion. The origin of AmMRJPs was recently examined. AmMRJP families showed similarity with yellow protein of *Drosophila melanogaster* (Albert *et al.*, 1999a). All AmMRJP families shared the *N*-terminal hydrophobic sequences and were characterized as glycoproteins suggesting that these proteins are secretary proteins as the yellow protein. In addition, four cysteines are conserved for AmMRJPs and the yellow protein.

Two full lengths of *A. mellifera* apisimin homologues were obtained. This peptide was recently purified and characterized in *A. mellifera* (Bilikova *et al.*, 2002). Due to high level expression of its mRNA during the whole life span of the honeybee, it was proposed that apisimin might play an important physiological role in the honeybee colony. The *N*-ternimal sequence of *A. mellifera* apisimin was" KTSISVK", which was nearly identical to KTSISAK for apisimin in *A. cerana*. After alignment, four deduced amino acid residues (or 18 nucleotides) were different between these sequences. An equal length of a 234 bp ORF encoding 78 amino acid was observed..

ESTs representing glucose oxidase and α-glucosidase homologues were also isolated. These enzymes involve in carbohydrate metabolism. The α-glucosidase hydrolyzed sucrose to glucose and fructose while glucose oxidase is needed to convert glucose to gluconic acid and hydrogen peroxide which acts as antiseptic substance into honey. Expression of α-glucosidase was examined and found that this gene possibly expressed partially in *A. cerana* nurse bees and fully expressed in *A. cerana* foragers. Results were contradictory to expression patterns of α-glucosidase and glucose oxidase *A. mellifera* where these genes were specifically expressed in hypopharyngeal glands of the forager bee but not in nurse bees (Ohashi *et al.*, 1996; 1997; 1999).

Only one EST (1.52 %) clone was identified as a homologue of heat shock protein (hsp) 86 of *Mus musculus*. Hsps are important proteins widely contributed under various stress conditions of organisms. Severson *et al.* (1990) reported effects of heat stress on expression of hsps in *A. mellifera*. Expression of hsp70 and hsp82 kDa were greatly elevated in bees (aged 9 days) exposed to the temperature of 42 °C for 4 hours whereas only slight increased expression levels were found in bees directly taken from the colony.

4.4 Semi-quantification of the mRNA level using competitive PCR

Expression levels of different AmMRJPs cDNA in hypopharuyngeal glands of nurse bees were examined (Klaudiny *et al.*, 1994; Schmitzova *et al.*, 1998; Albert *et al.*, 1999b). AmMRJP3 and AmMRJP4 comprised 8 % and 2 % of total mRNA, respectively (Klaudiny *et al.*, 1994). Northern blot analysis showed that the ratio of AmMRJP1:AmMRJP2: AmMRJP3 mRNAs was 3:1:1 (Schmitzova *et al.*, 1998).

Initially, semi-quantitative PCR for estimating expression levels of AcMRJP genes in hypopharyngeal glands of *A. cerana* were attempted by endogenous comparison with the expression level of housekeeping genes (28S ribosomal and elongation factor 1αF2). Since expression of AcMRJPs gene was significantly higher than that of housekeeping genes, AcMRJPs were amplified and reached in the plateau of amplification much sooner than both 28S ribosomal and elongation factor 1αF2 (data not shown).

The detection approach was then changed to semi-quantitative PCR using competition between mRNA and its genomic DNA. The ratio of mRNA quantity between AcMRJP1:AcMRJP2:AcMRJP3 was 3.3 : 1.6 : 1 as nearly identical as that reported in *A. mellifera*. However, the ratio obtained from purified AcMRJPs was 7.2 : 2.7 : 1. Therefore, AcMRJP1 was predominated whereas AcMRJP3 appeared at the lowest amount at both transcriptional and translational level. Different ratio of AcMRJP families at transcriptional and translational levels implied translational regulation of these proteins.

The putative transcriptional factor binding site of each AcMRJP family was proposed (Malecova *et al.*, 2003). Genomic DNA of these genes revealed that

AmMRJPs are present as single-copy genes per haploid genome but different transcription factor binding sites for dead ringer (DRI) and ultraspiracle (USP) were found in regulatory sequences. DRI acts as an activator and repressor for transcriptional regulation of AmMRJPs while USP is one of hormone receptor family regulating cell differentiation and development. Both DRI and USP binding efficiency may cause different ratio between each family of AmMRJPs.

Due to age-dependent, expression of the gene encoding for a 64 kDa protein (identified as AmMRJP3) was specifically expressed in the hypopharyngeal gland of nurse bees whereas the gene encoding for a 56 kDa protein (identified as AmMRJP4) was expressed in both of nurse and forager bees (Ohashi *et al.*, 1997). Using microarray and northern blot analysis, the gene encoding for AmMRJP2 was expressed in the forager bees and over-expressed with caffeine treatment (Kucharski and Maleszka, 2002).

Expression of each RJ protein family at different developmental stages of *A. cerana* workers was examined in this study. AcMRJPs encoded genes were expressed in hypopharyngeal glands of nurse and forager bees, but not in newly emerged bee. Different levels of AcMRJPs expression were also found in those bees. The expression level of each family in nurse bee was greater than that in forager bees. Juvenile hormone (JH) may be directly regulated expression of these genes in nurse bees (Malecova *et al.*, 2003).